

**METHOD AND APPARATUS FOR PRESENTING IMAGES  
REPRESENTATIVE OF AN UTTERANCE WITH  
CORRESPONDING DECODED SPEECH**

**Abstract of the Disclosure**

5           Apparatus for presenting images representative of one or more words in an  
utterance with corresponding decoded speech includes, in one aspect, a visual detector for  
capturing images of body movements (e.g., lip and/or mouth movements) corresponding  
to the one or more words in the utterance coupled to a visual feature extractor. The visual  
feature extractor receives time information from an automatic speech recognition (ASR)  
10       system and operatively processes the captured images from the visual detector to generate  
one or more image segments based on the time information relating to one or more  
decoded words in the utterance, each image segment corresponding to a decoded word in  
the utterance. An image player coupled to the visual feature extractor presents an image  
segment with a corresponding decoded word. The image segment may be presented as an  
15       animation of successive images in time, whereby a user is provided multiple sources of  
information for comprehending the utterance and can more easily ascertain the  
relationship between the body movements and the corresponding decoded speech.

1500-193-APP